Fixed-Width Output Analysis for Markov Chain Monte Carlo

Galin L. Jones
School of Statistics
University of Minnesota
galin@stat.umn.edu

Murali Haran
Department of Statistics
The Pennsylvania State University
mharan@stat.psu.edu

Brian S. Caffo
Department of Biostatistics
Johns Hopkins University
bcaffo@jhsph.edu

Ronald Neath
School of Statistics
University of Minnesota
rneath@stat.umn.edu

February 9, 2005 (Revised October 5, 2005)

Abstract

Markov chain Monte Carlo is a method of producing a correlated sample in order to estimate features of a target distribution via ergodic averages. A fundamental question is when should sampling stop? That is, when are the ergodic averages good estimates of the desired quantities? We consider a method that stops the simulation when the width of a confidence interval based on an ergodic average is less than a user-specified value. Hence calculating a Monte Carlo standard error is a critical step in assessing the simulation output. We consider the regenerative simulation and batch means methods of estimating the

variance of the asymptotic normal distribution. We give sufficient conditions for the strong consistency of both methods and investigate their finite sample properties in a variety of examples.

1 Introduction

Suppose our goal is to calculate $E_\pi g := \int_{\mathsf{X}} g(x)\pi(dx)$ with π a probability distribution having support X and g a real-valued, π-integrable function. Also, suppose π is such that Markov chain Monte Carlo (MCMC) is the only viable method for estimating $E_\pi g$.

Let $X = \{X_0, X_1, X_2, \dots\}$ be a time-homogeneous, aperiodic, π-irreducible, positive Harris recurrent Markov chain with state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and invariant distribution π. (See Meyn and Tweedie (1993) for definitions.) In this case, we say that X is Harris ergodic and the Ergodic Theorem implies that, with probability 1,

$$\bar{g}_n := \frac{1}{n}\sum_{i=0}^{n-1} g(X_i) \to E_\pi g \quad \text{as } n \to \infty. \tag{1}$$

Given an MCMC algorithm that simulates X it is conceptually easy to generate large amounts of data and use $\bar{g}_n$ to obtain an arbitrarily precise estimate of $E_\pi g$.

There are several methods for deciding when n is sufficiently large; i.e., when to terminate the simulation. The simplest is to terminate the computation whenever patience runs out. This approach is unsatisfactory since the user would not have any idea about the accuracy of $\bar{g}_n$. Alternatively, with several preliminary (and necessarily short) runs the user might be able to make an informed guess about the variability in $\bar{g}_n$ and hence make an a priori choice of n. Another method would be to monitor the sequence of $\bar{g}_n$ until it appears to have stabilized. None of these methods are automated and hence are inefficient uses of user time and Monte Carlo resources. Moreover, they provide only a point estimate of $E_\pi g$ without additional work.

Convergence diagnostics are also sometimes used to terminate the simulation (Cowles and Carlin, 1996). Some convergence diagnostics are available in software, e.g. the R package `boa`,

and hence may be considered automated. However, none of the diagnostics of which we are aware explicitly address how well $\bar{g}_n$ estimates $E_\pi g$; this is discussed again in subsection 4.1.1.

An alternative is to calculate a Monte Carlo standard error and use it to terminate the simulation when the width of a confidence interval falls below a specified value. Under regularity conditions (see Section 2) the Markov chain X and function g will admit a central limit theorem (CLT); that is,

$$\sqrt{n}(\bar{g}_n - \mathrm{E}_\pi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2) \tag{2}$$

as $n \to \infty$ where $\sigma_g^2 := \mathrm{var}_\pi\{g(X_0)\} + 2\sum_{i=1}^{\infty} \mathrm{cov}_\pi\{g(X_0), g(X_i)\}$. Given an estimate of σ_g^2, say $\hat{\sigma}_n^2$, we can form a confidence interval for $\mathrm{E}_\pi g$. If this interval is too large then the value of n is increased and simulation continues until the interval is sufficiently small; this is a common way of choosing n (e.g., see Fishman, 1996; Geyer, 1992; Jones and Hobert, 2001). Notice that the final Monte Carlo sample size is random. We study sequential fixed-width methods which formalize this approach. In particular, the simulation terminates the first time

$$t_* \frac{\hat{\sigma}_n}{\sqrt{n}} + p(n) \le \epsilon \tag{3}$$

where t_* is an appropriate quantile, $p(n) \ge 0$ on $\mathbb{Z}_+$ and $\epsilon > 0$ is the desired half-width. The role of p is to ensure that the simulation is not terminated prematurely due to a poor estimate of σ_g^2. One possibility is to fix $n^* > 0$ and take $p(n) = \epsilon I(n \le n^*)$ where I is the usual indicator function.

Sequential statistical procedures have a long history; see Lai (2001) for an overview and commentary. Moreover, classical approaches to sequential fixed-width confidence intervals such as those found in Chow and Robbins (1965), Liu (1997) and Nadas (1969) are known to work well. However, the classical procedures are not relevant to the current work since they assume the observations are random samples.

In a simulation context, procedures based on (3) were studied most notably by

Glynn and Whitt (1992) who established that these procedures are *asymptotically valid* in that if our goal is to have a $100(1-\delta)\%$ confidence interval with width 2ϵ then

$$\Pr(\mathrm{E}_\pi g \in \mathrm{Int}[T(\epsilon)]) \to 1-\delta \quad \text{as } \epsilon \to 0 \tag{4}$$

where $T(\epsilon)$ is the first time that (3) is satisfied and $\mathrm{Int}[T(\epsilon)]$ is the interval at this time. Glynn and Whitt's conditions for asymptotic validity are substantial: (i) A functional central limit theorem (FCLT) holds; (ii) $\hat{\sigma}_n^2 \to \sigma_g^2$ with probability 1 as $n \to \infty$; and (iii) $p(n) = o(n^{-1/2})$. Markov chains frequently enjoy an FCLT under the same conditions that ensure a CLT. However, in the context of MCMC, little work has been done on establishing conditions for (ii) to hold. Thus one of our goals is to give conditions under which some common methods provide strongly consistent estimators of σ_g^2. Specifically, our conditions require the sampler to be either uniformly or geometrically ergodic. The MCMC community has expended considerable effort in establishing such mixing conditions for a variety of samplers; see Jones and Hobert (2001) and Roberts and Rosenthal (1998, 2004) for some references and discussion.

We consider two methods for estimating the variance of the asymptotic normal distribution, regenerative simulation (RS) and non-overlapping batch means (BM). Both have strengths and weaknesses; essentially, BM is easier to implement but RS is on a stronger theoretical footing. For example, when used with fixed number of batches BM *cannot* be even weakly consistent for σ_g^2. We give conditions for the consistency of RS and show that BM can provide a consistent estimation procedure by allowing the batch sizes to increase (in a specific way) as n increases. In this case it is denoted CBM to distinguish it from the standard fixed-batch size version which we denote BM. This was addressed by Damerdji (1994) but, while the approach is similar, our regularity conditions on X are weaker. Also, the regularity conditions required to obtain strong consistency of the batch means estimator are slightly stronger than those required by RS. Finally, it is important to note that RS and CBM do not require that X be stationary; hence burn-in is not required.

The justification of fixed-width methods is entirely asymptotic so it is not clear how the finite sample properties of BM, CBM, and RS compare in typical MCMC settings. For this reason, we conduct a simulation study in the context of two benchmark examples and two realistic examples, one of which is a complicated frequentist problem and one which involves a high-dimensional posterior. Roughly speaking, we find that BM performs poorly while RS and CBM are comparable.

The rest of this article is organized as follows. Section 2 fixes notation and contains a brief discussion of some relevant Markov chain theory. In Section 3 we consider RS and CBM. Then Section 4 contains the examples.

2 Basic Markov Chain Theory

For $n \in \mathbb{N} := \{1, 2, 3, \ldots\}$ let $P^n(x, dy)$ be the n-step Markov transition kernel; that is, for $x \in \mathsf{X}$ and $A \in \mathcal{B}(\mathsf{X})$, $P^n(x, A) = \Pr(X_n \in A | X_0 = x)$. A Harris ergodic Markov chain X enjoys a strong form of convergence. Specifically, if $\lambda(\cdot)$ is a probability measure on $\mathcal{B}(\mathsf{X})$ then

$$\|P^n(\lambda, \cdot) - \pi(\cdot)\| \downarrow 0 \quad \text{as } n \to \infty, \tag{5}$$

where $P^n(\lambda, A) := \int_{\mathsf{X}} P^n(x, A)\lambda(dx)$ and $\|\cdot\|$ is the total variation norm. Suppose there exists an extended real-valued function $M(x)$ and a nonnegative decreasing function $\kappa(n)$ on $\mathbb{Z}_+$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\kappa(n) \ . \tag{6}$$

When $\kappa(n) = t^n$ for some $t < 1$ say X is *geometrically ergodic* if M is unbounded and *uniformly ergodic* if M is bounded. *Polynomial ergodicity of order m* where $m \geq 0$ means M may be unbounded and $\kappa(n) = n^{-m}$.

Also, P satisfies *detailed balance* with respect to π if

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx) \qquad \text{for all } x, y \in \mathsf{X} \ . \tag{7}$$

Note that Metropolis-Hastings samplers satisfy (7) by construction but many Gibbs samplers do not. We are now in position to give conditions for the existence of a CLT.

Theorem. Let X be a Harris ergodic Markov chain on X with invariant distribution π and suppose $g : \mathsf{X} \to \mathbb{R}$ is a Borel function. Assume one of the following conditions:

1. X is polynomially ergodic of order $m > 1$, $\mathrm{E}_\pi M < \infty$ and there exists $B < \infty$ such that $|g(x)| < B$ almost surely;
2. X is polynomially ergodic of order m, $\mathrm{E}_\pi M < \infty$ and $\mathrm{E}_\pi |g(x)|^{2+\delta} < \infty$ for some $\delta > 0$ where $m\delta > 2 + \delta$;
3. X is geometrically ergodic and $\mathrm{E}_\pi[g^2(x)(\log^+ |g(x)|)] < \infty$;
4. X is geometrically ergodic, satisfies (7) and $\mathrm{E}_\pi g^2(x) < \infty$; or
5. X is uniformly ergodic and $\mathrm{E}_\pi g^2(x) < \infty$.

Then, for any initial distribution, as $n \to \infty$

$$\sqrt{n}(\bar{g}_n - \mathrm{E}_\pi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2) \ .$$

Remark 1. The theorem was proved by Ibragimov and Linnik (1971) (condition 5), Roberts and Rosenthal (1997) (condition 4), Doukhan et al. (1994) (condition 3). See Jones (2004) for details on conditions 1 and 2.

Remark 2. Conditions 3, 4 and 5 of the theorem are also sufficient to guarantee the existence of an FCLT; see Doukhan et al. (1994), Roberts and Rosenthal (1997) and Billingsley (1968), respectively.

Remark 3. The mixing conditions on the Markov chain X stated in Theorem 2 are not necessary for the CLT; see, for example, Chen (1999), Meyn and Tweedie (1993) and Nummelin (2002). However, the weaker conditions are often prohibitively difficult to check in situations where MCMC is appropriate.

Remark 4. There are constructive techniques for verifying the existence of an appropriate M and κ from (6) (Meyn and Tweedie, 1993, Ch. 15). For example, one method

of establishing geometric ergodicity requires finding a function $V : \mathsf{X} \to [1, \infty)$ and a small set $C \in \mathcal{B}(\mathsf{X})$ such that

$$PV(x) \leq \lambda V(x) + bI(x \in C) \quad \forall\, x \in \mathsf{X} \tag{8}$$

where $PV(x) := \int V(y)P(x, dy)$, $0 < \lambda < 1$ and $b < \infty$. Substantial effort has been devoted to establishing convergence rates for MCMC algorithms via (8) or related techniques. For example, Hobert and Geyer (1998), Hobert et al. (2002), Jones and Hobert (2004), Marchev and Hobert (2004), Mira and Tierney (2002), Robert (1995), Roberts and Polson (1994), Roberts and Rosenthal (1999), Rosenthal (1995, 1996) and Tierney (1994) examined Gibbs samplers while Christensen et al. (2001), Douc and Soulier (2004), Fort and Moulines (2000, 2003), Geyer (1999), Jarner and Hansen (2000), Jarner and Roberts (2002), Meyn and Tweedie (1994), and Mengersen and Tweedie (1996) analyzed Metropolis-Hastings algorithms.

2.1 The Split Chain

An object that is important to the study of both RS and CBM is the *split chain* $X' := \{(X_0, \delta_0), (X_1, \delta_1), (X_2, \delta_2), \ldots\}$ which has state space $\mathsf{X} \times \{0, 1\}$. The construction of X' requires a *minorization condition*; i.e., a function $s : \mathsf{X} \mapsto [0, 1]$ for which $E_\pi s > 0$ and a probability measure Q such that

$$P(x, A) \geq s(x)\, Q(A) \qquad \text{for all } x \in \mathsf{X} \text{ and } A \in \mathcal{B}(\mathsf{X})\ . \tag{9}$$

When X is countable it is easy to see that (9) holds by fixing $x_* \in \mathsf{X}$, setting $s(x) = I(x = x_*)$ and $Q(\cdot) = P(x_*, \cdot)$. Mykland et al. (1995) and Rosenthal (1995) give prescriptions that are often useful for establishing (9) in general spaces. Note that (9) allows us to write $P(x, dy)$ as a mixture of two distributions,

$$P(x, dy) = s(x)\, Q(dy) + [1 - s(x)]\, R(x, dy),$$

where $R(x, dy) := [1 - s(x)]^{-1} [P(x, dy) - s(x)\, Q(dy)]$ is the *residual* distribution (define $R(x, dy)$ as 0 if $s(x) = 1$). This mixture gives us a recipe for simulating X': given

$X_i = x$, generate $\delta_i \sim \text{Bernoulli}(s(x))$. If $\delta_i = 1$, then draw $X_{i+1} \sim Q(\cdot)$, else draw $X_{i+1} \sim R(x, \cdot)$.

The two chains, X and X' are closely related since X' will inherit properties such as aperiodicity and positive Harris recurrence and the sequence $\{X_i : i = 0, 1, \dots\}$ obtained from X' has the same transition probabilities as X. Also, X and X' converge to their respective stationary distributions at exactly the same rate.

If $\delta_i = 1$, then time $i + 1$ is a *regeneration time* when X' probabilistically restarts itself. Specifically, suppose we start X' with $X_0 \sim Q$. Then each time that $\delta_i = 1$, $X_{i+1} \sim Q$. Let $0 = \tau_0 < \tau_1 < \cdots$ be the regeneration times. That is, set $\tau_{r+1} = \min\{i > \tau_r : \delta_{i-1} = 1\}$. Also assume that X' is run for R tours; that is, the simulation is stopped the Rth time that a $\delta_i = 1$. Let τ_R denote the total length of the simulation and N_r be the length of the rth tour; that is, $N_r = \tau_r - \tau_{r-1}$. Define

$$S_r = \sum_{i=\tau_{r-1}}^{\tau_r - 1} g(X_i)$$

for $r = 1, \dots, R$. The (N_r, S_r) pairs are iid since each is based on a different tour. In the sequel we will make repeated use of the following lemma which generalizes Theorem 2 of Hobert et al. (2002).

Lemma 1. Let X be a Harris ergodic Markov chain with invariant distribution π. Assume that (9) holds and that X is geometrically ergodic. Let $p \geq 1$ be an integer.

1. If $E_\pi |g|^{2(p-1)+\delta} < \infty$ for some $\delta > 0$ then $E_Q N_1^p < \infty$ and $E_Q S_1^p < \infty$.
2. If $E_\pi |g|^{2p+\delta} < \infty$ for some $\delta > 0$ then $E_Q N_1^p < \infty$ and $E_Q S_1^{p+\delta} < \infty$.

Proof. See Appendix A. □

3 Output Analysis

3.1 Regenerative Simulation

Regenerative simulation is based on directly simulating the split chain. However, using the mixture approach described above is problematic since simulation from $R(x, dy)$ is challenging. Mykland et al. (1995) suggest a method for avoiding this issue. Suppose (9) holds and that the measures $P(x, \cdot)$ and $Q(\cdot)$ admit densities $k(\cdot|x)$ and $q(\cdot)$, respectively. Then the following recipe allows us to simulate X'. Assume $X_0 \sim q(\cdot)$; this is typically quite easy to do, see Mykland et al. (1995) for some examples. Also, note that this means burn-in is irrelevant. Draw $X_{i+1} \sim k(\cdot|x)$, that is, draw from the sampler at hand, and get δ_i by simulating from the distribution of $\delta_i|X_i, X_{i+1}$ with

$$\Pr(\delta_i = 1 \,|\, X_i, X_{i+1}) = \frac{s(X_i)q(X_{i+1})}{k(X_{i+1} \,|\, X_i)} \,. \tag{10}$$

Example 1. In a slight abuse of notation let π also denote the density of the target distribution. Consider an independence Metropolis-Hastings sampler with proposal density ν. This chain works as follows: Let the current state be $X_i = x$. Draw $y \sim \nu$ and independently draw $u \sim \text{Uniform}(0, 1)$. If

$$u < \frac{\pi(y)\nu(x)}{\pi(x)\nu(y)}$$

then set $X_{i+1} = y$ otherwise set $X_{i+1} = x$. Mykland et al. (1995) derive (10) for this case. Let $c > 0$ be a user-specified constant. Then conditional on an acceptance, i.e. $X_i = x$ and $X_{i+1} = y$

$$\Pr(\delta_i = 1 \,|\, X_i = x, X_{i+1} = y) = \begin{cases} c \max\left\{\frac{\nu(x)}{\pi(x)}, \frac{\nu(y)}{\pi(y)}\right\} & \text{if } \min\left\{\frac{\pi(x)}{\nu(x)}, \frac{\pi(y)}{\nu(y)}\right\} > c \\ \frac{1}{c} \max\left\{\frac{\pi(x)}{\nu(x)}, \frac{\pi(y)}{\nu(y)}\right\} & \text{if } \max\left\{\frac{\pi(x)}{\nu(x)}, \frac{\pi(y)}{\nu(y)}\right\} < c \\ 1 & \text{otherwise} \,. \end{cases} \tag{11}$$

Note that we do not need to know the normalizing constants of π or ν to calculate (11).

In discrete state spaces regenerations can be easy to identify. In particular, a regeneration occurs whenever the chain returns to any fixed state; for example, when the Metropolis-Hastings chain accepts a move to the fixed state. This regeneration scheme is most useful when the state space is not too large but potentially complicated; see subsection 4.3. It will not be useful when the state space is extremely large because returns to the fixed state are too infrequent. Further practical advice on implementing and automating RS is given in Brockwell and Kadane (2005), Gilks et al. (1998), Geyer and Thompson (1995), Hobert et al. (2002), Hobert et al. (2005) and Jones and Hobert (2001).

Implementation of RS is simple once we can effectively simulate the split chain. For example, the Ergodic Theorem implies that

$$\bar{g}_{\tau_R} = \frac{1}{\tau_R} \sum_{j=0}^{\tau_R - 1} g(X_j) \to \mathrm{E}_\pi g$$

with probability 1 as $R \to \infty$ and hence estimating $\mathrm{E}_\pi g$ is routine.

We now turn our attention to calculating a Monte Carlo standard error for $\bar{g}_{\tau_R}$. Let E_Q denote the expectation for the split chain started with $X_0 \sim Q(\cdot)$. Also, let $\bar{N}$ be the average tour length; that is, $\bar{N} = R^{-1} \sum_{r=1}^{R} N_r$. Since the (N_r, S_r) pairs are iid the strong law implies with probability 1, $\bar{N} \to E_Q N_1$ which is finite by positive recurrence. If $E_Q N_1^2 < \infty$ and $E_Q S_1^2 < \infty$ it follows that a CLT holds; i.e., as $R \to \infty$

$$\sqrt{R}(\bar{g}_{\tau_R} - \mathrm{E}_\pi g) \xrightarrow{d} \mathrm{N}\left(0, \xi_g^2\right) \tag{12}$$

where, as shown in Hobert et al. (2002), $\xi_g^2 = E_Q(S_1 - N_1 E_\pi g)^2 / (E_Q N_1)^2$. An obvious estimator of ξ_g^2 is

$$\hat{\xi}_{RS}^2 := \frac{1}{\bar{N}^2} \frac{1}{R} \sum_{r=1}^{R} (S_r - \bar{g}_{\tau_R} N_r)^2 \; .$$

Now consider

$$
\begin{aligned}
\hat{\xi}^2_{RS} - \xi^2_g &= \frac{1}{\bar{N}^2}\frac{1}{R}\sum_{r=1}^{R}(S_r - \bar{g}_{\tau_R}N_r)^2 - \frac{E_Q(S_1 - N_1E_\pi g)^2}{(E_QN_1)^2} \pm \frac{E_Q(S_1 - N_1E_\pi g)^2}{\bar{N}^2} \\
&= \frac{1}{\bar{N}^2}\frac{1}{R}\sum_{r=1}^{R}\left[(S_r - \bar{g}_{\tau_R}N_r)^2 - E_Q(S_1 - N_1E_\pi g)^2 \pm (S_r - N_rE_\pi g)^2\right] \\
&\qquad + \left[E_Q(S_1 - N_1E_\pi g)^2\left(\frac{1}{\bar{N}^2} - \frac{1}{E_QN_1^2}\right)\right] .
\end{aligned}
$$

Using this representation and repeated application of the strong law shows that $\hat{\xi}^2_{RS} - \xi^2_g \to 0$ with probability 1 as $R \to \infty$ (also see Hobert et al., 2002). It is typically difficult to check that $E_QN_1^2 < \infty$ and $E_QS_1^2 < \infty$. However, using Lemma 1 yields the following result.

Proposition. Let X be a Harris ergodic Markov chain with invariant distribution π. Assume that $E_\pi|g|^{2+\delta} < \infty$ for some $\delta > 0$, (9) holds and that X is geometrically ergodic. Then (12) holds and $\hat{\xi}^2_{RS} \to \xi^2_g$ w. p. 1 as $R \to \infty$.

Fix $\epsilon > 0$ and let z denote an appropriate standard normal quantile. An asymptotically valid fixed-width procedure results by terminating the simulation the first time

$$
z\,\frac{\hat{\xi}_{RS}}{\sqrt{R}} + p(R) \le \epsilon \; . \tag{13}
$$

3.2 Batch Means

In standard batch means the output of the sampler is broken into batches of equal size that are assumed to be approximately independent. (This is not strictly necessary; c.f., the method of overlapping batch means.) Suppose the algorithm is run for a total of $n = ab$ iterations (hence $a = a_n$ and $b = b_n$ are implicit functions of n) and define

$$
\bar{Y}_j := \frac{1}{b}\sum_{i=(j-1)b}^{jb-1} g(X_i) \qquad \text{for } j = 1, \ldots, a \; .
$$

The batch means estimate of σ^2_g is

$$
\hat{\sigma}^2_{BM} = \frac{b}{a-1}\sum_{j=1}^{a}(\bar{Y}_j - \bar{g}_n)^2 \; . \tag{14}
$$

With a fixed number of batches (14) is not a consistent estimator of σ_g^2 (Glynn and Iglehart, 1990; Glynn and Whitt, 1991). On the other hand, if the batch size *and* the number of batches are allowed to increase as the overall length of the simulation does it may be possible to obtain consistency. The first result in this direction is due to Damerdji (1994) which we now describe. The major assumption made by Damerdji (1994) is the existence of a strong invariance principle. Let $B = \{B(t), t \geq 0\}$ denote a standard Brownian motion. A strong invariance principle holds if there exists a nonnegative increasing function $\gamma(n)$ on the positive integers, a constant $0 < \sigma_g < \infty$ and a sufficiently rich probability space such that

$$\left| \sum_{i=1}^{n} g(X_i) - n\mathrm{E}_\pi g - \sigma_g B(n) \right| = O(\gamma(n)) \quad \text{w.p. 1 as } \; n \to \infty \tag{15}$$

where the w.p. 1 in (15) means for almost all sample paths. In particular, Damerdji (1994) assumed (15) held with $\gamma(n) = n^{1/2-\alpha}$ where $0 < \alpha \leq 1/2$. However, it would seem a daunting task to directly check this condition in any given application. In an attempt to somewhat alleviate this difficulty we have the following lemma.

Lemma 2. Let $g : \mathsf{X} \to \mathbb{R}$ be a Borel function and let X be a Harris ergodic Markov chain with invariant distribution π.

1. If X is uniformly ergodic and $\mathrm{E}_\pi|g|^{2+\delta} < \infty$ for some $\delta > 0$ then (15) holds with $\gamma(n) = n^{1/2-\alpha}$ where $\alpha < \delta/(24 + 12\delta)$.
2. If X is geometrically ergodic, (9) holds and $\mathrm{E}_\pi|g|^{4+\delta} < \infty$ for some $\delta > 0$ then (15) holds with $\gamma(n) = n^\alpha \log n$ where $\alpha = 1/(2 + \delta)$.

Proof. The first part of the lemma is an immediate consequence of Theorem 4.1 of Philipp and Stout (1975) and the fact that uniformly ergodic Markov chains enjoy exponentially fast uniform mixing. The second part follows from our Lemma 1 and Theorem 2.1 in Csáki and Csörgő (1995). $\square$

Using part 1 of Lemma 2 we can state Damerdji's result as follows.

Proposition. (Damerdji, 1994) Assume $g : \mathsf{X} \to \mathbb{R}$ such that $\mathrm{E}_\pi |g|^{2+\delta} < \infty$ for some $\delta > 0$ and let X be a Harris ergodic Markov chain with invariant distribution π. Further, suppose X is uniformly ergodic. If

1. $a_n \to \infty$ as $n \to \infty$,

2. $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$,

3. $b_n^{-1} n^{1-2\alpha} \log n \to 0$ as $n \to \infty$ where $\alpha \in (0, \delta/(24 + 12\delta))$ and

4. there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$

then as $n \to \infty$, $\hat{\sigma}^2_{BM} \to \sigma^2_g$ w. p. 1.

In Appendix B we use part 2 of Lemma 2 to extend Proposition 3.2 to geometrically ergodic Markov chains.

Proposition. Assume $g : \mathsf{X} \to \mathbb{R}$ such that $\mathrm{E}_\pi |g|^{4+\delta} < \infty$ for some $\delta > 0$ and let X be a Harris ergodic Markov chain with invariant distribution π. Further, suppose X is geometrically ergodic. If

1. $a_n \to \infty$ as $n \to \infty$,

2. $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$,

3. $b_n^{-1} n^{2\alpha} [\log n]^3 \to 0$ as $n \to \infty$ where $\alpha = 1/(2 + \delta)$ and

4. there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$

then as $n \to \infty$, $\hat{\sigma}^2_{BM} \to \sigma^2_g$ w. p. 1.

Remark 5. There is no assumption of stationarity in Propositions 3.2 or 3.2. Hence burn-in is not required to implement CBM.

Remark 6. Consider using $b_n = \lfloor n^\theta \rfloor$ and $a_n = \lfloor n/b_n \rfloor$. Proposition 3.2 requires that $1 > \theta > 1 - 2\alpha > 1 - \delta/(12 + 6\delta) > 5/6$ but Proposition 3.2 requires only $1 > \theta > (1 + \delta/2)^{-1} > 0$.

Under the conditions of Propositions 3.2 or 3.2 an asymptotically valid fixed-width procedure for estimating $E_\pi g$ results if we terminate the simulation the first time

$$t_{a_n - 1} \frac{\hat{\sigma}_{BM}}{\sqrt{n}} + p(n) \leq \epsilon$$

where t_{a_n-1} is the appropriate quantile from a student's t distribution with $a_n - 1$ degrees of freedom.

3.3 Practical Implementation Issues

Making practical use of the preceding theory requires (i) a moment condition; (ii) establishing geometric ergodicity of the sampler at hand; (iii) choosing $p(n)$; (iv) using RS requires (9) or at least (10); and (v) CBM requires choosing a_n and b_n.

Since a moment condition is required even in the iid case we do not view (i) as restrictive. Consider (ii). It is easy to construct examples where the convergence rate is so slow that a Markov chain CLT does not hold (Roberts, 1999) so the importance of establishing the rate of convergence in (6) should not be underestimated. On the other hand, the MCMC community has expended considerable effort in trying to understand when certain Markov chains are geometrically ergodic; see the references in Remark 4. In our view, this is not the obstacle that it once was.

Regarding (iii), we know of no work on choosing an optimal $p(n)$. Recall that the theory requires $p(n) = o(n^{-1/2})$. In our examples we use $p(n) = \epsilon I(n \leq n^*)$ where $n^* > 0$ is fixed. Since n^* is typically chosen based on empirical experience with the sampler at hand we might want a penalty for sample sizes greater than n^* so another reasonable choice might be $p(n) = \epsilon I(n \leq n^*) + Cn^{-k}$ for some $k > 1/2$ and $C > 0$.

The issue in (iv), i.e., calculating (9) or (10) is commonly viewed as overly burdensome. However, in our experience, this calculation need not be troublesome. For example, Mykland et al. (1995) give recipes for constructing (9) and (10) for Metropolis-Hastings independence and random walk samplers; recall (11). There is also some work

on establishing these conditions for very general models; see Hobert et al. (2005). Finally, Brockwell and Kadane (2005) and Geyer and Thompson (1995) have shown that regenerations can be made to occur naturally via simulated tempering.

Consider (v). As we noted in Remark 6, it is common to choose the batch sizes according to $b_n = \lfloor n^\theta \rfloor$ for some θ. Song and Schmeiser (1995) and Chien (1988) have addressed the issue of what value of θ should be used from different theoretical points of view. In particular, Chien (1988) showed that (under regularity conditions) using $\theta = 1/2$ results in the batch means approaching asymptotic normality at the fastest rate. Song and Schmeiser (1995) showed that (under different regularity conditions) using $\theta = 1/3$ minimizes the asymptotic mean-squared error of $\hat{\sigma}^2_{BM}$. Note that Remark 6 shows that $\theta = 1/3$ requires a stronger moment condition than $\theta = 1/2$. We further address this issue in Section 4.

3.4 Alternatives to BM and RS

We chose to focus on BM and RS since in MCMC settings they seem to be the most common methods for estimating the variance of the asymptotic normal distribution. However, there are other methods which may enjoy strong consistency; e.g. see Damerdji (1991), Geyer (1992), Nummelin (2002) and Peligrad and Shao (1995). In particular, Damerdji (1991) uses a strong invariance principle to obtain strong consistency of certain spectral variance estimators under conditions similar to those required in Proposition 3.2. Apparently, this can be extended to geometrically ergodic chains via Lemma 2 to obtain a result with regularity conditions similar to Proposition 3.2. However, we do not pursue this further here.

4 Examples

In this section we investigate the finite sample performance of RS, BM with 30 batches, and CBM with $b_n = \lfloor n^{1/3} \rfloor$ and $b_n = \lfloor n^{1/2} \rfloor$ in four examples. In particular, we examine the coverage probabilities and half-widths of the resulting intervals as well as the required simulation effort. While each example concerns a different statistical model and MCMC sampler there are some commonalities. In each case we perform many independent replications of the given MCMC sampler. The number of replications ranges from 2000 to 9000 depending on the complexity of the example. We used all methods on the *same* output from each replication of the MCMC sampler. When the half-width of a 95% interval with $p(n) = \epsilon I(n \geq n^*)$ (or $p(R) = \epsilon I(R \geq R^*)$ for RS) is less than ϵ for a particular method, that procedure was stopped and the chain length recorded. Our choice of n^* is different for each example and was chosen based on our empirical experience with the given Markov chain. Other procedures would continue until all of them were below the targeted half-width, at which time a single replication was complete. In order to estimate the coverage probabilities we need true values of the quantities of interest. These are not analytically available in three of our examples. Our solution is to obtain precise estimates of the truth through independent methods which are different for each example. The details are described below. The results are reported in Table 2.

4.1 Toy Example

Consider estimating the mean of a Pareto(α, β) distribution, i.e., $\alpha\beta/(\beta-1)$, $\beta > 1$, using a Metropolis-Hastings independence sampler with a Pareto(α, λ) candidate. Let π be the target density and ν be the proposal density. Assume $\beta \geq \lambda$. Then for $x \geq \alpha$

$$\frac{\pi(x)}{\nu(x)} = \frac{\beta}{\lambda} \alpha^{\beta-\lambda} x^{\lambda-\beta} \leq \frac{\beta}{\lambda} \, .$$

By Theorem 2.1 in Mengersen and Tweedie (1996) this sampler is uniformly ergodic and

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq \left(1 - \frac{\lambda}{\beta}\right)^n .$$

In order to ensure the moment conditions required for Proposition 3.2 we set $\beta = 10$ and $\lambda = 9$ in which case the right hand side is 10^{-n}. Hence this sampler converges extremely fast. Implementation of RS was accomplished using (11) with $c = 1.5$.

4.1.1 Comparing convergence diagnostics with CBM

As noted by a referee, one method for terminating the simulation is via convergence diagnostics. Consider the method of Geweke (1992) which is a diagnostic that seems close in spirit to the current work. Geweke's diagnostic (GD) is based on a Markov chain CLT and hence does not apply much more generally than CBM; the same can be said for many other diagnostics. GD uses a hypothesis test to ascertain when $\bar{g}_n$ has stabilized.

In the remainder of this subsection we compare GD and CBM in terms of mean-squared error (MSE) and chain length. To this end we ran 9000 independent replications of the independence sampler with $\alpha = 1$, $\beta = 10$ and $\lambda = 9$. We used CBM and GD on the output in the following manner. For each replication we set $n^* = 45$ but the R package `boa` required a minimum of 120 iterations in order to calculate GD. After the minimum was achieved and the cutoff for a particular method was attained we noted the chain length and the current estimate of $E_\pi g$. The cutoff for CBM was to set the desired half-width to $\epsilon = .005$. The result of using GD is a p-value. We chose four values (.05, .10, .2 and .4) for the threshold in an attempt to tune the computation. The results are reported in Table 1. As we previously noted, this sampler mixes extremely well. Thus it is not surprising that using GD results in a small estimated MSE. However, using CBM results in much smaller MSE than GD. The average chain lengths make it is clear that GD stops the simulation much too soon. Moreover, changing the p-value

Method	Cutoff	Estimated MSE	Average Chain Length
CBM ($b_n = \lfloor n^{1/3} \rfloor$)	$\epsilon = .005$	$6.65 \times 10^{-6} (9.9 \times 10^{-8})$	2428 (5)
CBM ($b_n = \lfloor n^{1/2} \rfloor$)	$\epsilon = .005$	$7.34 \times 10^{-6} (1.2 \times 10^{-8})$	2615 (3)
Geweke	p-value=.4	$1.17 \times 10^{-4} (2 \times 10^{-6})$	202.6 (3.4)
Geweke	p-value=.2	$1.30 \times 10^{-4} (2 \times 10^{-6})$	148.9 (1.6)
Geweke	p-value=.1	$1.34 \times 10^{-4} (2 \times 10^{-6})$	133.4 (.9)
Geweke	p-value=.05	$1.37 \times 10^{-4} (2 \times 10^{-6})$	127.4 (.5)

Table 1: Summary statistics for CBM versus GD for Example 4.1. Standard errors of estimates are in parentheses.

threshold for GD does not result in substantial improvements in estimation accuracy.

4.2 A Hierarchical Model

Efron and Morris (1975) present a data set that gives the raw batting averages (based on 45 official at-bats) and a transformation ($\sqrt{45} \arcsin(2x - 1)$) for 18 Major League Baseball players during the 1970 season. Rosenthal (1996) considers the following conditionally independent hierarchical model for the transformed data. Suppose for $i = 1, \ldots, K$ that

$$\begin{aligned} Y_i|\theta_i &\sim \mathrm{N}(\theta_i, 1) \qquad & \theta_i|\mu, \lambda &\sim \mathrm{N}(\mu, \lambda) \\ \lambda &\sim \mathrm{IG}(2, 2) \qquad & f(\mu) &\propto 1 \; . \end{aligned} \tag{16}$$

(Note that we say $W \sim \mathrm{Gamma}(\alpha, \beta)$ if its density is proportional to $w^{\alpha-1}e^{-\beta w} I(w > 0)$ and if $X \sim \mathrm{Gamma}(b, c)$ then $X^{-1} \sim \mathrm{IG}(b, c)$.) Rosenthal (1996) introduces a Harris ergodic block Gibbs sampler that has the posterior, $\pi(\theta, \mu, \lambda|y)$, characterized by the hierarchy in (16) as its invariant distribution. This Gibbs sampler completes a one-step transition $(\lambda', \mu', \theta') \to (\lambda, \mu, \theta)$ by drawing from the distributions of $\lambda|\theta'$ then $\mu|\theta', \lambda$ and subsequently $\theta|\mu, \lambda$. The full conditionals needed to implement this sampler are

given by

$$\lambda|\theta, y \sim \text{IG}\left(2 + \frac{K-1}{2}, 2 + \frac{\sum(\theta_i - \bar{\theta})^2}{2}\right), \quad \mu|\theta, \lambda, y \sim \text{N}\left(\bar{\theta}, \frac{\lambda}{K}\right),$$

$$\theta_i|\lambda, \mu, y \overset{\text{ind}}{\sim} \text{N}\left(\frac{\lambda y_i + \mu}{\lambda + 1}, \frac{\lambda}{\lambda + 1}\right).$$

Rosenthal proved geometric ergodicity of the associated Markov chain. However, MCMC is not required to sample from the posterior; in Appendix C we develop an accept-reject sampler that produces an iid sample from the posterior. Also in Appendix C we derive an expression for the probability of regeneration (10).

We focus on estimating the posterior mean of θ_9, the "true" long-run (transformed) batting average of the Chicago Cubs' Ron Santo. It is straightforward to check that the moment conditions for CBM and RS are met. Finally, we employed our accept-reject sampling algorithm to generate 9×10^7 independent draws from $\pi(\theta_9|y)$ which were then used to estimate the posterior mean of θ_9 which we assumed to be the truth.

4.3 Calculating Exact Conditional P-Values

Agresti (2002, p. 432) reports data that correspond to pairs of scorings of tumor ratings by two pathologists. A linear by linear association model specifies that the log of the Poisson mean in cell i, j satisfies

$$\log \mu_{ij} = \alpha + \beta_i + \gamma_j + \delta\, ij \ .$$

A parameter free null distribution for testing goodness-of-fit is obtained by conditioning on the sufficient statistics for the parameters, i.e., the margins of the table and $\sum_{ij} n_{ij}\, ij$, where the n_{ij} are the observed cell counts. The resulting conditional distribution is a generalization of the hypergeometric distribution. An exact p-value for goodness-of-fit versus a saturated alternative can be calculated by summing the conditional probabilities of all tables satisfying the margins and the additional constraint and having deviance statistics larger than the observed.

For the current data set there are over twelve billion tables that satisfy the margin constraints but an exhaustive search revealed that there are only roughly 34,000 tables that also satisfy the constraint induced by $\sum_{ij} n_{ij}\, ij$. We will denote this set of permissible tables by Γ. Now the desired p-value is given by

$$\sum_{y \in \Gamma} I[d(y) \geq d(y_{obs})]\, \pi(y) \tag{17}$$

where $d(\cdot)$ is the deviance function and π denotes the generalized hypergeometric. Since we have enumerated Γ we find that the true exact p-value is .044 whereas the chi-squared approximation yields a p-value of .368. However, a different data set with different values of the sufficient statistics will have a different reference set which must be enumerated in order to find the exact p-value. This would be too computationally burdensome to implement generally and hence it is common to resort to MCMC-based approximations (see e.g. Caffo and Booth, 2001; Diaconis and Sturmfels, 1998).

To estimate (17) we will use the Metropolis-Hastings algorithm developed in Caffo and Booth (2001). This algorithm is also employed by the R package `exactLoglinTest`. The associated Markov chain is Harris ergodic and its invariant distribution is the appropriate generalized hypergeometric distribution. Moreover, the chain is uniformly ergodic and since we are estimating the expectation of a bounded function the regularity conditions for both RS and CBM are easily met.

Implementation of RS is straightforward. As we mentioned earlier, in finite state spaces regenerations occur whenever the chain returns to any fixed state. In order to choose the fixed state we ran the algorithm for 1000 iterations and chose the state which had the highest probability with respect to the stationary distribution. The same fixed state was used in each replication.

4.4 A Model-Based Spatial Statistics Application

Consider the Scottish lip cancer data set (Clayton and Kaldor, 1987) which consists of the number of cases of lip cancer registered in each of the 56 (pre-reorganization) counties of Scotland, together with the expected number of cases given the age-sex structure of the population. We assume a Poisson likelihood for areal (spatially aggregated) data. Specifically, for $i = 1, ..., N$ we assume that given μ_i the disease counts Y_i are conditionally independent and

$$Y_i|\mu_i \sim \text{Poisson}(E_i e^{\mu_i}) \tag{18}$$

where E_i is the known 'expected' number of disease events in the ith region assuming constant risk and μ_i is the log-relative risk of disease for the ith region. Set $\phi = (\phi_1, \ldots, \phi_N)^T$. Each μ_i is modeled as $\mu_i = \theta_i + \phi_i$ where

$$\theta_i|\tau_h \sim \ \text{N}(0, 1/\tau_h), \qquad \phi|\tau_c \sim \ \text{CAR}(\tau_c) \ \propto \ \tau_c^{N/2} \exp\left(-\frac{\tau_c}{2}\phi^T Q\phi\right), \text{ and}$$

$$Q_{ij} = \begin{cases} n_i & \text{if } i = j \\ 0 & \text{if } i \text{ is not adjacent to } j \\ -1 & \text{if } i \text{ is adjacent to } j \end{cases}$$

with n_i the number of neighbors for the ith region. Each θ_i captures the ith region's extra-Poisson variability due to area-wide heterogeneity, while each ϕ_i captures the ith region's excess variability attributable to regional clustering. The priors on the precision parameters are $\tau_h \sim$ Gamma(1, .01) and $\tau_c \sim$ Gamma(1, .02). This is a challenging model to consider since the random effects parameters (θ_i, ϕ_i) are not identified in the likelihood, and the spatial prior used is improper. Also, no closed form expressions are available for the marginal distributions of the parameters, and the posterior distribution has $2N + 2$ dimensions (114 for the lip cancer data).

Haran and Tierney (2004) establish uniform ergodicity of a Harris ergodic Metropolis-Hastings independence sampler with invariant distribution $\pi(\theta, \phi, \tau_h, \tau_c|y)$ where $\theta = (\theta_1, \ldots, \theta_N)^T$ and a heavy-tailed proposal. In our implementation of RS we used the

formula for the probability of a regeneration given by (11) with $\log c = -342.72$. Using the empirical supremum of the ratio of the invariant density to the proposal density (based on several draws from the proposal) guided the choice of c.

We focus on estimating the posterior expectation of ϕ_7, the log-relative risk of disease for County 7 attributable to spatial clustering. Finally, we used an independent run of length 10^7 to obtain an estimate which we treated as the 'true value'.

4.5 Summary

Table 2 reveals that the estimates of the coverage probabilities are all less than the desired .95. However, examining the standard errors shows that only BM is significantly less in all of the examples and the estimated coverage probability for RS is *not* significantly different from .95 in 3 out of 4. The story for CBM is more complicated in that the coverage depends on the choice of b_n. Using $b_n = \lfloor n^{1/3} \rfloor$ gives the best coverage for the examples in Sections 4.1 and 4.2 while $b_n = \lfloor n^{1/2} \rfloor$ is superior for those in Sections 4.3 and 4.4. The reason for this is that the Markov chains in Sections 4.1 and 4.2 mix exceptionally well and hence smaller batch sizes can be tolerated. However, the examples in Sections 4.3 and 4.4 are realistic problems and hence the chains do not mix as well so that larger batch sizes are required. Thus we would generally recommend using $b_n = \lfloor n^{1/2} \rfloor$.

The example in subsection 4.3 deserves to be singled out due to the low estimated coverage probabilities. The goal in this example was to estimate a fairly small probability, a situation in which the Wald interval is known to have poor coverage even in iid settings.

While RS and CBM appear comparable in terms of coverage probability RS tends to result in slightly longer runs than CBM which in turn results in longer runs than BM. Moreover, RS and CBM are comparable in their ability to produce intervals that meet the target half-width more closely than BM. Also, the intervals for RS are apparently

more stable than those of CBM and BM. Finally, BM underestimates the Monte Carlo standard error and therefore suggests stopping the chain too early.

While RS has a slight theoretical advantage over CBM their finite sample properties appear comparable. Also, like RS, CBM avoids the burn-in issue, which has been a long standing obstacle to MCMC practitioners. In addition, CBM enjoys the advantage of being slightly easier to implement. Thus CBM clearly has a place in the tool kit of MCMC practitioners.

A Proof of Lemma 1

A.1 Preliminary Results

Recall the split chain X' and that $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ denote the regeneration times; i.e., $\tau_{r+1} = \min\{i > \tau_r : \delta_{i-1} = 1\}$.

Lemma 3. (Hobert et al., 2002, Lemma 1) Let X be a Harris ergodic Markov chain and assume that (9) holds. Then for any function $h : \mathsf{X}^\infty \to \mathbb{R}$

$$\mathrm{E}_\pi |h(X_0, X_1, \ldots)| \geq c\mathrm{E}_Q |h(X_0, X_1, \ldots)|$$

where $c = \mathrm{E}_\pi s$.

Lemma 4. (Hobert et al., 2002, Lemma 2) Let X be a Harris ergodic Markov chain and assume that (9) holds. If X is geometrically ergodic, then there exists a $\beta > 1$ such that $\mathrm{E}_\pi \beta^{\tau_1} < \infty$.

Corollary 1. Assume the conditions of Lemma 4. For any $a > 0$

$$\sum_{i=0}^{\infty} [\Pr_\pi(\tau_1 \geq i+1)]^a \leq (\mathrm{E}_\pi \beta^{\tau_1})^a \sum_{i=0}^{\infty} \beta^{-a(i+1)} < \infty \ .$$

A.2 Proof of Lemma 1

We prove only part 2 of the lemma as part 1 is similar. Without loss of generality we assume $0 < \delta < 1$. By Lemma 3, it is enough to verify that $\mathrm{E}_\pi \tau_1^p < \infty$ and $\mathrm{E}_\pi S_1^{p+\delta} < \infty$. Lemma 4 shows that $\mathrm{E}_\pi \tau_1^p < \infty$ for any $p > 0$. Note that

$$\left(\sum_{i=0}^{\tau_1-1} g(X_i)\right)^{p+\delta} \leq \left(\sum_{i=0}^{\tau_1-1} |g(X_i)|\right)^{p+\delta} = \left(\sum_{i=0}^{\infty} I(0 \leq i \leq \tau_1 - 1)|g(X_i)|\right)^{p+\delta}$$
$$\leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \sum_{i_{p+1}=0}^{\infty} \left[\prod_{j=1}^{p} I(0 \leq i_j \leq \tau_1 - 1)|g(X_{i_j})|\right] I(0 \leq i_{p+1} \leq \tau_1 - 1)|g(X_{i_{p+1}})|^{\delta}$$

and hence

$$\mathrm{E}_\pi S_1^{p+\delta} \leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \sum_{i_{p+1}=0}^{\infty} \mathrm{E}_\pi \left(\left[\prod_{j=1}^{p+1} I(0 \leq i_j \leq \tau_1 - 1)\right]\left[\prod_{j=1}^{p} |g(X_{i_j})|\right] |g(X_{i_{p+1}})|^{\delta}\right)$$
$$\leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \sum_{i_{p+1}=0}^{\infty} \left[\mathrm{E}_\pi I(0 \leq i_1 \leq \tau_1 - 1)|g(X_{i_1})|^2\right]^{1/2} \times$$
$$\cdots \times \left[\mathrm{E}_\pi I(0 \leq i_p \leq \tau_1 - 1)|g(X_{i_p})|^{2^p}\right]^{1/2^p} \left[\mathrm{E}_\pi I(0 \leq i_{p+1} \leq \tau_1 - 1)|g(X_{i_{p+1}})|^{2^p\delta}\right]^{1/2^p}$$

where the second inequality follows with repeated application of Cauchy-Schwartz. Set $a_j = 1 + 2^j/\delta$ and $b_j = 1 + \delta/2^j$ for $j = 1, 2, \ldots, p$ and apply Hölder's inequality to obtain

$$\mathrm{E}_\pi I(0 \leq i_j \leq \tau_1 - 1)|g(X_{i_j})|^{2^j} \leq [\mathrm{E}_\pi I(0 \leq i_j \leq \tau_1 - 1)]^{1/a_j} \left[\mathrm{E}_\pi |g(X_{i_j})|^{2^j+\delta}\right]^{1/b_j} .$$

Note that

$$c_j := \left[\left(\mathrm{E}_\pi |g(X_{i_j})|^{2^j+\delta}\right)^{1/b_j}\right]^{1/2^p} < \infty .$$

Also, if $a_{p+1} = 1 + 2^p$ and $b_{p+1} = 1 + 1/2^p$ then

$$\mathrm{E}_\pi I(0 \leq i_{p+1} \leq \tau_1 - 1)|g(X_{i_{p+1}})|^{2^p\delta} \leq [\mathrm{E}_\pi I(0 \leq i_{p+1} \leq \tau_1 - 1)]^{\frac{1}{a_{p+1}}} \left[\mathrm{E}_\pi |g(X_{i_{p+1}})|^{\delta(2^p+\delta)}\right]^{\frac{1}{b_{p+1}}} .$$

Notice that

$$c_{p+1} := \left[\left(\mathrm{E}_\pi |g(X_{i_{p+1}})|^{\delta(2^p+\delta)}\right)^{1/b_j}\right]^{1/2^p} < \infty$$

and set $c = \max\{c_1, \ldots, c_{p+1}\}$. Then an appeal to Corollary 1 yields

$$\mathrm{E}_\pi S_1^{p+\delta} \leq c \left[\prod_{j=1}^{p} \sum_{i_j=0}^{\infty} \{\Pr_\pi(\tau_1 \geq i_j + 1)\}^{1/(a_j 2^j)}\right] \left[\sum_{i_{p+1}=0}^{\infty} \{\Pr_\pi(\tau_1 \geq i_j + 1)\}^{1/(a_{p+1}2^p)}\right] < \infty .$$

B Proof of Proposition 3.2

B.1 Preliminary Results

Recall that $B = \{B(t), t \geq 0\}$ denotes a standard Brownian motion. Define

$$\tilde{\sigma}_*^2 = \frac{b_n}{a_n - 1} \sum_{j=0}^{a_n - 1} \left(\bar{B}_j(b_n) - \bar{B}(n)\right)^2 \tag{19}$$

where $\bar{B}_j(b_n) = b_n^{-1}\left(B((j+1)b_n) - B(jb_n)\right)$ and $\bar{B}(n) = n^{-1}B(n)$.

Lemma 5. (Damerdji, 1994, p. 508) For all $\epsilon > 0$ and for almost all sample paths there exists $n_0(\epsilon)$ such that for all $n \geq n_0$ (Damerdji, 1994, p. 508)

$$|\bar{B}_j(b_n)| \leq \sqrt{2}(1+\epsilon) b_n^{-1/2} [\log(n/b_n) + \log\log n]^{1/2} \; . \tag{20}$$

Lemma 6. (Csörgő and Révész, 1981) For all $\epsilon > 0$ and for almost all sample paths there exists $n_0(\epsilon)$ such that for all $n \geq n_0$

$$|B(n)| < (1+\epsilon)[2n \log\log n]^{1/2} \; . \tag{21}$$

B.2 Proof of Proposition 3.2

Proposition 3.2 follows from Lemma 2 and the following two lemmas:

Lemma 7. (Damerdji, 1994, Proposition 3.1) Assume

1. $b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$ and
2. there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$

then as $n \to \infty$, $\tilde{\sigma}_*^2 \to 1$ a.s.

Lemma 8. Assume that (15) holds with $\gamma(n) = n^{\alpha} \log n$ where $\alpha = 1/(2+\delta)$. If

1. $a_n \to \infty$ as $n \to \infty$,
2. $b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$ and
3. $b_n^{-1} n^{2\alpha} [\log n]^3 \to 0$ as $n \to \infty$ where $\alpha = 1/(2+\delta)$

then as $n \to \infty$, $\hat{\sigma}^2_{BM} - \sigma_g^2 \tilde{\sigma}_*^2 \to 0$ a.s.

Proof of Lemma 8. Recall that $X = \{X_1, X_2, \ldots\}$ is a Harris ergodic Markov chain. Define the process Y by $Y_i = g(X_i) - \mathrm{E}_\pi g$ for $i = 1, 2, 3, \ldots$. Then

$$\hat{\sigma}^2_{BM} = \frac{b_n}{a_n - 1} \sum_{j=0}^{a_n - 1} \left(\bar{Y}_j(b_n) - \bar{Y}(n)\right)^2$$

where $\bar{Y}_j(b_n) = b_n^{-1} \sum_{i=1}^{b_n} Y_{jb_n+i}$ for $j = 0, \ldots, a_n - 1$ and $\bar{Y}(n) = n^{-1} \sum_{i=1}^{n} Y_i$. Since

$$\bar{Y}_j(b_n) - \bar{Y}(n) = \bar{Y}_j(b_n) - \bar{Y}(n) \pm \sigma_g \bar{B}_j(b_n) \pm \sigma_g \bar{B}(n)$$

we have

$$\begin{aligned}
\left|\hat{\sigma}^2_{BM} - \sigma_g^2 \tilde{\sigma}_*^2\right| \leq \frac{b_n}{a_n - 1} \sum_{j=0}^{a_n - 1} & \left[(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))^2 + (\bar{Y}(n) - \sigma_g \bar{B}(n))^2\right. \\
& + |2(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))(\bar{Y}(n) - \sigma_g \bar{B}(n))| + |2\sigma_g(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))\bar{B}_j(b_n)| \\
& + |2\sigma_g(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))\bar{B}(n)| + |2\sigma_g(\bar{Y}(n) - \sigma_g \bar{B}(n))\bar{B}_j(b_n)| \\
& \left. + |2\sigma_g(\bar{Y}(n) - \sigma_g \bar{B}(n))\bar{B}(n)|\right] .
\end{aligned}$$

Now we will consider each term in the sum and show that it tends to 0.

1. Our assumptions say that there exists a constant C such that for all large n

$$\left|\sum_{i=1}^{n} g(X_i) - n\mathrm{E}_\pi g - \sigma_g B(n)\right| < C n^\alpha \log n \quad a.s. \tag{22}$$

   Note that

$$\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n) = \frac{1}{b_n}\left[\sum_{i=1}^{(j+1)b_n} Y_i - \sigma_g B((j+1)b_n)\right] - \frac{1}{b_n}\left[\sum_{i=1}^{jb_n} Y_i - \sigma_g B(jb_n)\right]$$

   and hence by (22)

$$\begin{aligned}
|\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n)| &\leq \frac{1}{b_n}\left[|\sum_{i=1}^{(j+1)b_n} Y_i - \sigma_g B((j+1)b_n)| + |\sum_{i=1}^{jb_n} Y_i - \sigma_g B(jb_n)|\right] \\
&< \frac{2}{b_n} C n^\alpha \log n
\end{aligned} \tag{23}$$

Then

$$\frac{b_n}{a_n - 1}\sum_{j=0}^{a_n-1}(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))^2 < 4C^2 \frac{a_n}{a_n - 1} b_n^{-1} n^{2\alpha} (\log n)^2 \ \to 0$$

as $n \to \infty$ by conditions 1 and 3.

2. Apply (22) to obtain

$$|\bar{Y}(n) - \sigma_g \bar{B}(n)| = n^{-1} |\sum_{i=1}^{n} Y_i - \sigma_g B(n)| < C n^{\alpha - 1} \log n \ . \tag{24}$$

Then

$$\frac{b_n}{a_n - 1}\sum_{j=0}^{a_n-1}(\bar{Y}(n) - \sigma_g \bar{B}(n))^2 < C^2 \frac{a_n}{a_n - 1} \frac{b_n}{n} \frac{(\log n)^2}{n^{1-2\alpha}} \to 0$$

as $n \to \infty$ by conditions 1 and 2 and since $1 - 2\alpha > 0$.

3. By (23) and (24)

$$|2(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))(\bar{Y}(n) - \sigma_g \bar{B}(n))| < 4C^2 b_n^{-1} n^{2\alpha - 1} (\log n)^2 \ .$$

Thus

$$\frac{b_n}{a_n - 1}\sum_{j=0}^{a_n-1} |2(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))(\bar{Y}(n) - \sigma_g \bar{B}(n))| < 4C^2 \frac{a_n}{a_n - 1} \frac{(\log n)^2}{n^{1-2\alpha}} \to 0$$

as $n \to \infty$ by condition 1 and since $1 - 2\alpha > 0$.

4. Since $b_n \geq 2$, (20) and (23) together imply

$$\begin{aligned}|(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))\bar{B}_j(b_n)| < 2^{3/2} C(1+\epsilon) b_n^{-1} \Big[& b_n^{-1} n^{2\alpha} (\log n)^2 \log(n/b_n) \\ & + b_n^{-1} n^{2\alpha} (\log n)^2 \log\log n \Big]^{1/2}\end{aligned}$$

Hence

$$\begin{aligned}\frac{b_n}{a_n - 1}\sum_{j=0}^{a_n-1} |2\sigma_g(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))\bar{B}_j(b_n)| \leq 8\sigma_g C(1+\epsilon)\frac{a_n}{a_n - 1} \Big[& b_n^{-1} n^{2\alpha} (\log n)^2 \log(n/b_n) \\ & + b_n^{-1} n^{2\alpha} (\log n)^2 \log\log n \Big]^{1/2} \ \to 0\end{aligned}$$

as $n \to \infty$ by conditions 1 and 3.

5. By (23) and (21) $|(\bar{Y}_j(b_n)-\sigma_g \bar{B}_j(b_n))\bar{B}(n)| < 4C(1+\epsilon)b_n^{-1}n^{-1/2+\alpha}(\log n)(\log\log n)^{1/2}$ so that

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}|2\sigma_g(\bar{Y}_j(b_n)-\sigma_g \bar{B}_j(b_n))\bar{B}(n)| < 8\sigma_g C(1+\epsilon)\frac{a_n}{a_n-1}\frac{(\log n)(\log\log n)^{1/2}}{n^{1/2-\alpha}} \to 0$$

as $n\to\infty$ by condition 1 and since $1/2-\alpha>0$.

6. Use (20) and (24) to get

$$|(\bar{Y}(n)-\sigma_g\bar{B}(n))\bar{B}_j(b_n)| < \sqrt{2}C(1+\epsilon)\frac{n^{\alpha-1}\log n}{\sqrt{b_n}}\left[\log(n/b_n)+\log\log n\right]^{1/2}$$

and hence using conditions 1, 2 and 3 shows that as $n\to\infty$

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}|2\sigma_g(\bar{Y}(n)-\sigma_g\bar{B}(n))\bar{B}_j(b_n)| <$$
$$4\sigma_g C(1+\epsilon)\frac{a_n}{a_n-1}\frac{b_n}{n}\left[b_n^{-1}n^{2\alpha}((\log n)^2\log(n/b_n)+(\log n)^2\log\log n)\right]^{1/2} \to 0$$

7. Now (21) and (24) imply $|(\bar{Y}(n)-\sigma_g\bar{B}(n))\bar{B}(n)| < 2C(1+\epsilon)n^{-3/2+\alpha}(\log n)^{3/2}$. Hence

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}|2\sigma_g(\bar{Y}(n)-\sigma_g\bar{B}(n))\bar{B}(n)| < 4\sigma_g C(1+\epsilon)\frac{a_n}{a_n-1}\frac{b_n}{n}\frac{(\log n)^{3/2}}{n^{1/2-\alpha}} \to 0$$

as $n\to\infty$ by conditions 1 and 2 and since $1/2-\alpha>0$.

C Calculations for Example 4.2

We consider a slightly more general formulation of the model given in (16). Suppose for $i=1,\ldots,K$ that

$$\begin{aligned} Y_i|\theta_i &\sim \mathrm{N}(\theta_i,a) & \theta_i|\mu,\lambda &\sim \mathrm{N}(\mu,\lambda) \\ \lambda &\sim \mathrm{IG}(b,c) & f(\mu) &\propto 1\,. \end{aligned} \tag{25}$$

where a,b,c are all known positive constants.

C.1 Sampling from $\pi(\theta, \mu, \lambda|y)$

Let $\pi(\theta, \mu, \lambda|y)$ be the posterior distribution corresponding to the hierarchy in (25). Note that θ is a vector containing all of the θ_i and that y is a vector containing all of the data. Consider the factorization

$$\pi(\theta, \mu, \lambda|y) = \pi(\theta|\mu, \lambda, y)\pi(\mu|\lambda, y)\pi(\lambda|y). \tag{26}$$

If it is possible to sequentially simulate from each of the densities on the right-hand side of (26) we can produce iid draws from the posterior. Now $\pi(\theta|\mu, \lambda, y)$ is the product of independent univariate normal densities, i.e. $\theta_i|\mu, \lambda, y \sim \mathrm{N}((\lambda y_i + a\mu)/(\lambda + a),\ a\lambda/(\lambda + a))$. Also, $\pi(\mu|\lambda, y)$ is a normal distribution, i.e. $\mu|\lambda, y \sim \mathrm{N}(\bar{y}, (\lambda + a)/K)$. Next

$$\pi(\lambda|y) \propto \frac{1}{\lambda^{b+1}(\lambda + a)^{(K-1)/2}} e^{-c/\lambda - s^2/2(\lambda+a)}$$

where $\bar{y} = K^{-1}\sum_{i=1}^{K} y_i$ and $s^2 = \sum_{i=1}^{K}(y_i - \bar{y})^2$. An accept-reject sampler with an $\mathrm{IG}(b, c)$ candidate can be used to sample from $\pi(\lambda|y)$ since if we let $g(\lambda)$ be the kernel of an $\mathrm{IG}(b, c)$ density

$$\sup_{\lambda \geq 0} \frac{1}{g(\lambda)\lambda^{b+1}(\lambda + a)^{(K-1)/2}} e^{-c/\lambda - s^2/2(\lambda+a)} = \sup_{\lambda \geq 0} (\lambda + a)^{(1-K)/2} e^{-s^2/2(\lambda+a)} = M < \infty$$

It is easy to show that the only critical point is $\hat{\lambda} = s^2/(K-1) - a$ which is where the maximum occurs if $\hat{\lambda} > 0$. But if $\hat{\lambda} \leq 0$ then the maximum occurs at 0.

C.2 Implementing regenerative simulation

We begin by establishing the minorization condition (9) for Rosenthal's (1996) block Gibbs sampler. For the one-step transition $(\lambda', \mu', \theta') \to (\lambda, \mu, \theta)$ the Markov transition density, p, is given by $p(\lambda, \mu, \theta|\lambda', \mu', \theta') = f(\lambda, \mu|\theta')f(\theta|\lambda, \mu)$. Note that $\mathsf{X} = \mathbb{R}^+ \times \mathbb{R}^1 \times$

$\mathbb{R}^K$. Fix a point $(\tilde{\lambda}, \tilde{\mu}, \tilde{\theta}) \in \mathsf{X}$ and let $D \subseteq \mathsf{X}$. Then

$$\begin{aligned} p(\lambda, \mu, \theta | \lambda', \mu', \theta') &= f(\lambda, \mu | \theta') f(\theta | \lambda, \mu) \\ &\geq f(\lambda, \mu | \theta') f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}} \\ &= \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}} \\ &\geq \left\{ \inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} \right\} f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}} \end{aligned}$$

and hence (9) will follow by setting

$$\varepsilon = \int_D f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) \, d\lambda \, d\mu \, d\theta,$$

$$s(\lambda', \mu', \theta') = \varepsilon \inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} \quad \text{and} \quad q(\lambda, \mu, \theta) = \varepsilon^{-1} f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}}.$$

Now using (10) shows that when $(\lambda, \mu, \theta) \in D$ the probability of regeneration is given by

$$\Pr(\delta = 1 | \lambda', \mu', \theta', \lambda, \mu, \theta) = \left\{ \inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} \right\} \frac{f(\lambda, \mu | \tilde{\theta})}{f(\lambda, \mu | \theta')} \tag{27}$$

Thus we need to calculate the infimum and plug into (27). To this end let $0 < d_1 < d_2 < \infty$, $-\infty < d_3 < d_4 < \infty$ and set $D = [d_1, d_2] \times [d_3, d_4] \times \mathbb{R}^K$. Define $V(\theta, \mu) = \sum_{i=1}^K (\theta_i - \mu)^2$ and note that

$$\inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} = \inf_{\lambda \in [d_1, d_2], \; \mu \in [d_3, d_4]} \exp\left\{ \frac{V(\tilde{\theta}, \mu) - V(\theta', \mu)}{2\lambda} \right\} = \exp\left\{ \frac{V(\tilde{\theta}, \hat{\mu}) - V(\theta', \hat{\mu})}{2\hat{\lambda}} \right\}$$

where $\hat{\mu} = d_4 I(\bar{\theta}' \leq \bar{\tilde{\theta}}) + d_3 I(\bar{\theta}' > \bar{\tilde{\theta}})$ and $\hat{\lambda} = d_2 I(V(\theta', \hat{\mu}) \leq V(\tilde{\theta}, \hat{\mu})) + d_1 I(V(\theta', \hat{\mu}) > V(\tilde{\theta}, \hat{\mu}))$. We find the fixed point with a preliminary estimate of the mean of the stationary distribution, and D to be centered at that point. Let $(\tilde{\lambda}, \tilde{\mu}, \tilde{\theta})$ be the ergodic mean for a preliminary Gibbs sampler run, and let S_λ and S_μ denote the usual sample standard deviations of λ and μ respectively. After some trial and error we took $d_1 = \max\left\{.01, \tilde{\lambda} - .5 S_\lambda\right\}$, $d_2 = \tilde{\lambda} + .5 S_\lambda$, $d_3 = \tilde{\mu} - S_\mu$ and $d_4 = \tilde{\mu} + S_\mu$.

Acknowledgments

The authors are grateful to Ansu Chatterjee, Jeff Rosenthal and Bill Sudderth for helpful conversations about this paper.

References

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Brockwell, A. E. and Kadane, J. B. (2005). Identification of regeneration times in MCMC simulation, with application to adaptive schemes. *Journal of Computational and Graphical Statistics*, 14:436–458.

Caffo, B. S. and Booth, J. G. (2001). A Markov chain Monte Carlo algorithm for approximating exact conditional probabilities. *Journal of Compuatational and Graphical Statistics*, 10:730–745.

Chen, X. (1999). Limit theorems for functionals of ergodic Markov chains with general state space. *Memoirs of the American Mathematical Society*, 139.

Chien, C.-H. (1988). Small sample theory for steady state confidence intervals. In Abrams, M., Haigh, P., and Comfort, J., editors, *Proceedings of the Winter Simulation Conference*, pages 408–413.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *The Annals of Mathematical Statistics*, 36:457–462.

Christensen, O. F., Moller, J., and Waagepetersen, R. P. (2001). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models. *Methodology and Computing in Applied Probability*, 3:309–327.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904.

Csáki, E. and Csörgő, M. (1995). On additive functionals of Markov chains. *Journal of Theoretical Probability*, 8:905–919.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press.

Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37:1424–1440.

Damerdji, H. (1994). Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research*, 19:494–512.

Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26:363–397.

Douc, R., F. G. M. E. and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *The Annals of Applied Probability*, 14:1353–1377.

Doukhan, P., Massart, P., and Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Annales de l'Institut Henri Poincare, Section B, Calcul des Probabilities et Statistique*, 30:63–82.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70:311–319.

Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York.

Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statistics and Probability Letters*, 49:401–410.

Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and their Applications*, 103:57–99.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, pages 169–188. Clarendon Press.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science*, 7:473–511.

Geyer, C. J. (1999). Likelihood inference for spatial point processes. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, pages 79–140. Chapman & Hall/CRC, Boca Raton.

Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90:909–920.

Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association*, 93:1045–1054.

Glynn, P. W. and Iglehart, D. L. (1990). Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15:1–16.

Glynn, P. W. and Whitt, W. (1991). Estimating the asymptotic variance with batch means. *Operations Research Letters*, 10:431–435.

Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2:180–198.

Haran, M. and Tierney, L. (2004). Perfect sampling for a Bayesian spatial model. Technical report, Pennsylvania State University, Department of Statistics.

Hobert, J. P. and Geyer, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, 67:414–430.

Hobert, J. P., Jones, G. L., Presnell, B., and Rosenthal, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, 89:731–743.

Hobert, J. P., Jones, G. L., and Robert, C. P. (2005). Using a Markov chain to construct a tractable approximation of an intractable probability distribution. *Scandinavian Journal of Statistics,* to appear.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Walters-Noordhoff, The Netherlands.

Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications*, 85:341–361.

Jarner, S. F. and Roberts, G. O. (2002). Polynomial convergence rates of Markov chains. *Annals of Applied Probability*, 12:224–247.

Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320.

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16:312–334.

Jones, G. L. and Hobert, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32:784–817.

Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, 11:303–351.

Liu, W. (1997). Improving the fully sequential sampling scheme of Anscombe-Chow-Robbins. *The Annals of Statistics*, 25:2164–2171.

Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *Journal of the American Statistical Association*, 99:228–238.

Mengersen, K. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24:101–121.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

Meyn, S. P. and Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, 4:981–1011.

Mira, A. and Tierney, L. (2002). Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics*, 29:1–12.

Mykland, P., Tierney, L., and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90:233–241.

Nadas, A. (1969). An extension of a theorem of Chow and Robbins on sequential confidence intervals for the mean. *The Annals of Mathematical Statistics*, 40:667–671.

Nummelin, E. (2002). MC's for MCMC'ists. *International Statistical Review*, 70:215–240.

Peligrad, M. and Shao, Q.-M. (1995). Estimation of the variance of partial sums for ρ-mixing random variables. *Journal of Multivariate Analysis*, 52:140–157.

Philipp, W. and Stout, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Memoirs of the American Mathematical Society*, 2:1–140.

Robert, C. P. (1995). Convergence control methods for Markov chain Monte Carlo algorithms. *Statistical Science*, 10:231–253.

Roberts, G. O. (1999). A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. *Journal of Applied Probability*, 36:1210–1217.

Roberts, G. O. and Polson, N. G. (1994). On the geometric convergence of the Gibbs sampler. *Journal of the Royal Statistical Society,* Series B, 56:377–384.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25.

Roberts, G. O. and Rosenthal, J. S. (1998). Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion). *Canadian Journal of Statistics*, 26:5–31.

Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society,* Series B, 61:643–660.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90:558–566.

Rosenthal, J. S. (1996). Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Statistics and Computing*, 6:269–275.

Song, W. T. and Schmeiser, B. W. (1995). Optimal mean-squared-error batch sizes. *Management Science*, 41:110–123.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22:1701–1762.

Example Section	Method	b_n	n^* / R^*	Average half width	Average Chain Length	Coverage Probability
4.1	CBM	$\lfloor n^{1/2} \rfloor$	45	.0048 (1.9×10^{-6})	2428 (5)	.923 (.003)
9000 reps	CBM	$\lfloor n^{1/3} \rfloor$	45	.0049 (8.0×10^{-7})	2615 (3)	.943 (.002)
$\epsilon = .005$	BM	$\lfloor n/30 \rfloor$	45	.0047 (2.4×10^{-6})	2342 (6)	.908 (.003)
	RS	-	30	.0049 (4.0×10^{-7})	2653 (2)	.948 (.002)
4.2	CBM	$\lfloor n^{1/2} \rfloor$	2000	.0194 (7.2×10^{-6})	5549 (13)	.930 (.004)
5000 reps	CBM	$\lfloor n^{1/3} \rfloor$	2000	.0198 (3.3×10^{-6})	5778 (6)	.947 (.003)
$\epsilon = .02$	BM	$\lfloor n/30 \rfloor$	2000	.0191 (1.1×10^{-5})	5279 (18)	.915 (.004)
	RS	-	50	.0198 (2.3×10^{-6})	5818 (12)	.945 (.003)
4.3	CBM	$\lfloor n^{1/2} \rfloor$	4000	.0049 (1.6×10^{-6})	56258 (405)	.920 (.006)
2000 reps	CBM	$\lfloor n^{1/3} \rfloor$	4000	.0049 (1.8×10^{-6})	46011 (499)	.869 (.008)
$\epsilon = .005$	BM	$\lfloor n/30 \rfloor$	4000	.0049 (1.7×10^{-6})	45768 (478)	.874 (.007)
	RS	-	20	.0049 (4.3×10^{-6})	58265 (642)	.894 (.007)
4.4	CBM	$\lfloor n^{1/2} \rfloor$	10000	.00396 (8.0×10^{-7})	168197 (270)	.934 (.005)
2000 reps	CBM	$\lfloor n^{1/3} \rfloor$	10000	.00398 (4.0×10^{-7})	137119 (125)	.900 (.006)
$\epsilon = .004$	BM	$\lfloor n/30 \rfloor$	10000	.00394 (1.2×10^{-6})	132099 (809)	.880 (.007)
	RS	-	25	.00398 (2.0×10^{-7})	179338 (407)	.942 (.005)

Table 2: Summary statistics for BM, CBM and RS. Standard errors of estimates are in parentheses.